medRxiv preprint doi: https://doi.org/10.1101/2022.05.27.22275037; this version posted June 7, 2022. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted medRxiv a license to display the preprint in perpetuity. It is made available under a CC-BY-NC-ND 4.0 International license .

Corticosteroids in COVID-19: Optimizing Observational Research through Target Trial Emulations

Katherine L. Hoffman, MS[1]
Edward J. Schenck, MD, MS[2]
Michael J. Satlin, MD[3]
William Whalen, MD, MS[2],
Di Pan, MD, MS[2]
Nicholas Williams, MPH[4]
Iván Díaz, PhD[1]

[1]Division of Biostatistics, Department of Population Health Sciences, Weill Cornell Medicine.
[2]Division of Pulmonary and Critical Care, Department of Medicine, Weill Cornell Medicine.
[3]Division of Infectious Disease, Department of Medicine, Weill Cornell Medicine.
[4]Department of Epidemiology, Mailman School of Public Health, Columbia University.

June 7, 2022

NOTE: This preprint reports new research that has not been certified by peer review and should not be used to guide clinical practice.

medRxiv preprint doi: https://doi.org/10.1101/2022.05.27.22275037; this version posted June 7, 2022. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted medRxiv a license to display the preprint in perpetuity. It is made available under a CC-BY-NC-ND 4.0 International license .

Abstract

Background: Observational research provides a unique opportunity to learn causal effects when randomized trials are unavailable, but obtaining the correct estimates hinges on a multitude of design and analysis choices. We illustrate the advantages of modern causal inference methods and compare to standard research practice to estimate the effect of corticosteroids on mortality in hospitalized COVID-19 patients in an observational dataset. We use several large RCTs to benchmark our results.

Methods: Our retrospective data consists of 3,298 COVID-19 patients hospitalized at New York-Presbyterian March 1-May 15, 2020. We design our study using the target trial framework. We estimate the effect of an intervention consisting of six days of corticosteroids administered at the time of severe hypoxia and contrast with an intervention consisting of no corticosteroids. The dataset includes dozens of time-varying confounders. We estimate the causal effects using a doubly robust estimator where the probabilities of treatment, outcome, and censoring are estimated using flexible regressions via super learning. We compare these analyses to standard practice in clinical research, consisting of two approaches: (i)Cox models for an exposure of corticosteroids receipt within various time windows of hypoxia, and (ii)Cox time-varying model where the exposure is daily administration of corticosteroids beginning at hospitalization.

Results: Our target trial emulation estimates corticosteroids to reduce 28-day mortality from 32% (95% confidence interval: 31-34) to 23% (21-24). This is qualitatively identical to the WHO's RCT meta-analysis result. Hazard ratios from the Cox models range in size and direction from 0.50 (0.41-0.62) to 1.08 (0.80-1.47) and all study designs suffer from various forms of bias.

Conclusion: We demonstrate that clinical research based on observational data can unveil true causal relations. However, the correctness of these effect estimates requires designing and analyzing the data based on principles which are different from the current standard in clinical research. Widespread communication and adoption of these design and analytical techniques is of high importance for the improvement of clinical research.

medRxiv preprint doi: https://doi.org/10.1101/2022.05.27.22275037; this version posted June 7, 2022. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted medRxiv a license to display the preprint in perpetuity. It is made available under a CC-BY-NC-ND 4.0 International license .

Introduction

Observational databases are invaluable resources when randomized controlled trials (RCTs) are infeasible or unavailable. However, the correctness of the conclusions gleaned from analyses of observational data hinges on the careful consideration of study design principles and choice of estimation methodology.

Threats to the validity of causality are pervasive in the clinical literature(1–4). A major reason for the failure to address these biases is the widespread adoption of a *model-first* approach to observational research. In this approach, a model is first chosen according to the data type and outcome of interest, and the quantity used to answer the research question is automatically determined by the model choice. For example, when faced with a time-to-event outcome, researchers automatically resort to a Cox regression model. It is common practice to then use the coefficients of the model or transformations thereof (e.g., hazard ratios) as the answer to the clinical question of interest.

A model-first approach induces multiple problems for the estimation of causal effects(5). First, model parameters often do not represent quantities of scientific interest or well-defined causal effects(6)). Second, assumptions such as the proportional hazards assumption used in Cox models are rarely correct in medical research since hazards cannot be proportional when a treatment effect changes over time(7). Third, regression models cannot correctly handle time-dependent feedback between confounders, treatment, and the outcome(1). Fourth, the model-first approach yields a tendency to interpret all coefficients in the model, which is a mistake known as the Table 2 fallacy(8). Lastly, model-first analyses often employ less-than-optimal model selection techniques, which may lead to improper variance estimates and model misspecification bias(9).

Recent developments in the causal inference literature provide researchers with a number of tools to alleviate the aforementioned biases. Frameworks such as the target trial emulation(10) and roadmap for causal inference(11) allow researchers to proceed with a *question-first* approach. Instead of defaulting to effect measures provided by regression models, a question-first approach begins by defining a hypothetical target trial and subsequent target of inference that answers the scientific question of interest. This is the so-called estimand, or quantity to be estimated. After

medRxiv preprint doi: https://doi.org/10.1101/2022.05.27.22275037; this version posted June 7, 2022. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted medRxiv a license to display the preprint in perpetuity. It is made available under a CC-BY-NC-ND 4.0 International license .

the estimand is chosen, the most appropriate statistical technique may be selected. Incorporating these principles can help clarify the research question, determine study eligibility requirements, identify enrollment and follow-up times, decide whether sufficient confounder data are available, and more(12,13). A question-first approach also allows researchers the freedom to select an estimation technique which mitigates model misspecification biases and increases the likelihood of obtaining a correct estimate.

We hypothesize that a question-first approach will have improved success in recovering causal effects as opposed to a model-first approach. To test this hypothesis, we use a retrospective cohort of 3,298 coronavirus 2019 (COVID-19) patients hospitalized at New-York-Presbyterian Hospital (NYPH) March 1-May 15, 2020. Lack of guidance for clinical practice at the beginning of the pandemic meant that high variability existed in the administration and timing of corticosteroids. While provider practice variability aids in the estimation of causal effects, the resulting complex longitudinal treatment patterns can complicate study design and analytical methods. Our dataset together with results from numerous RCTs on corticosteroids provide a unique opportunity to assess various design and analysis methods. We benchmark the results of our analyses against the effect measures obtained in the World Health Organization (WHO)'s RCT meta-analysis(14).

Methods

Hypothetical target trial

Population

Inclusion criteria is all adult patients with COVID-19 who were admitted to NYPH Weill Cornell, Lower Manhattan Hospital, or NYPH Queens. Cases are confirmed through reverse-transcriptase–polymerase chain-reaction assays performed on nasopharyngeal swab specimens. Patients who have chronic use of corticosteroids prior to hospitalization or who are transferred into the hospital from an outside hospital are excluded.

medRxiv preprint doi: https://doi.org/10.1101/2022.05.27.22275037; this version posted June 7, 2022. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted medRxiv a license to display the preprint in perpetuity. It is made available under a CC-BY-NC-ND 4.0 International license .

Intervention strategies

Patients are randomized on their first day of hospitalization to receive either (1)standard of care therapy (without corticosteroids) or (2)standard of care plus a corticosteroid regimen to be administered if and when criteria for severe hypoxia are met. The corticosteroid dosage is a minimum of 0.5 mg/kg body weight of methylprednisolone equivalent per 24-hour period and the duration of therapy is six days(15). Corticosteroids include prednisone, prednisolone, methylprednisolone, hydrocortisone, and dexamethasone and choice of drug is at the attending physician's discretion. Severe hypoxia is defined as the initiation of high flow nasal cannula, venti-mask, noninvasive or invasive mechanical ventilation, or an oxygen saturation of <93% after the patient is on 6 Liters of supplemental oxygen via nasal cannula.

Outcome and estimand

The primary outcome is 28-day mortality from time of randomization. The estimand of interest is the difference in 28-day mortality rates between the two treatment strategies.

Data analysis plan

In this hypothetical target trial with no loss-to-follow-up, we analyze the difference in proportion of patients who experienced the outcome between those who were randomized to the "standard of care" treatment regime and those who were randomized to the "standard of care plus corticosteroids at time of hypoxia" treatment regime.

Emulation using observational data

Cohort description and data source

The target trial emulation uses retrospective data from patients who meet the hypothetical trial's eligibility criteria between March 1 and May 15, 2020. Demographics, comorbidity, intubation, death, and discharge data were manually abstracted by trained medical professionals into a secure REDCap database(16). These were supplemented with a data repository housing laboratory, procedure, diagnosis, medication, and flowsheet data documented as part of standard care(17). Patients are followed for 28 days from hospitalization and lost to follow-up by discharge or transfer to an external hospital system.

medRxiv preprint doi: https://doi.org/10.1101/2022.05.27.22275037; this version posted June 7, 2022. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted medRxiv a license to display the preprint in perpetuity. It is made available under a CC-BY-NC-ND 4.0 International license .

Treatment strategies and measurement

To emulate the target trial corticosteroid treatment strategy, we estimate the effect of a hypothetical dynamic treatment regime(18), whereby each patient is administered six days of corticosteroids if and when they meet severe hypoxia criteria. This dynamic regime is contrasted with a static regime where patients never receive corticosteroids.

We measure severe hypoxia using vital signs (for oxygen saturation) and flowsheet data (for supplemental oxygen) and define it in the same way as our target trial. We measure corticosteroid exposure using the Medical Administration Record. We compute cumulative mg/kg dosing of corticosteroids over rolling 24-hour windows, and if a patient received >0.5 mg/kg methylprednisolone equivalent, they are denoted as having corticosteroids exposure that day.

Unlike our target trial, patients in the observational study are subject to loss-to-follow-up. Thus, our emulation requires a hypothetical intervention whereby patients are not loss to follow-up, so that we can observe their 28-day mortality status. A sample of observed data is shown in Supplemental Figure 1. An illustration of the treatment regimes as they relate to the observed data are shown in Figure 1.

Confounders

In contrast to the hypothetical trial, treatment assignment in the observational study is not randomized and depends on physiological characteristics of each patient. Correct emulation requires: (i) careful consideration of all possible confounders, and (ii) careful adjustment for these confounders in data analysis.

Baseline confounders include age, sex, race, ethnicity, Body Mass Index (BMI), comorbidities (coronary artery disease (CAD), cerebral vascular event, hypertension, diabetes mellitus (DM), cirrhosis, chronic obstructive pulmonary disease, active cancer, asthma, interstitial lung disease, chronic kidney disease (CKD/ESRD), immunosuppression, human immunodeficiency virus (HIV)-infection, home oxygen use), mode of respiratory support within three hours of hospital arrival, and hospital admission location.

Time-dependent confounders include heart rate, pulse oximetry percentage, respiratory rate, temperature, systolic/diastolic blood pressure, blood urea nitrogen (BUN)-creatinine ratio, creatinine, neutrophils, lymphocytes, platelets, bilirubin, blood glucose, D-dimers, C-reactive

medRxiv preprint doi: https://doi.org/10.1101/2022.05.27.22275037; this version posted June 7, 2022. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted medRxiv a license to display the preprint in perpetuity. It is made available under a CC-BY-NC-ND 4.0 International license .

protein, activated partial thromboplastin time, prothrombin time, arterial partial pressures of oxygen and carbon dioxide, and level of supplemental oxygen support. Figure 2 summarizes the relationship between confounders, treatment, and outcomes in the form of a Directed Acyclic Graph (DAG).

Outcome and estimand

Our estimand of interest is the difference in 28-day mortality rates in a hypothetical world where we had implemented the two different corticosteroid treatment strategies, as well as an intervention to prevent loss-to-follow-up. Under the assumption that treatment and loss-to-follow-up each day are randomized conditional on baseline and time-dependent confounders, this estimand is identified by the longitudinal g-computation formula(19). This longitudinal g-computation formula for our two corticosteroids treatment regimens with a censoring intervention will be our estimand of interest.

Data analysis plan

Correct emulation of a target trial requires proper adjustment for measured confounding through estimation of the g-computation formula. It is important to use estimation methods capable of fitting the data using flexible mathematical relationships so that confounding is appropriately removed, especially when the number of baseline and time-dependent confounders is large.

Several methods can be used to estimate the g-computation formula (e.g., inverse probability weighting (IPW), parametric g-formula, targeted minimum loss-based estimators (TMLE), sequentially doubly robust estimators (SDR), etc.)(20,21). These estimation methods rely on two kinds of mathematical models: (i)models of the outcome as a function of the time-dependent confounders, and (ii)models of treatment as a function of time-dependent confounders. Methods that use only one of these models are often called *singly robust*, because their correctness relies on the ability to correctly specify one of the models (e.g., IPW relies on estimating treatment models correctly). Estimation methods that use both of these models are often called *doubly robust*, because they remain correct under misspecification of one of the two models.

Furthermore, doubly robust estimators such as TMLE and SDR allow the use of machine learning to flexibly fit relevant treatment and outcome regressions(22,23). This is desirable because these regression functions might include complex relationships between exposures and

medRxiv preprint doi: https://doi.org/10.1101/2022.05.27.22275037; this version posted June 7, 2022. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted medRxiv a license to display the preprint in perpetuity. It is made available under a CC-BY-NC-ND 4.0 International license .

treatments, and capturing those relationships is not possible using simpler models such as the Cox proportional hazards(24).

The primary analysis is conducted using SDR estimation with a dynamic intervention, time-varying confounders, and a time-to-event outcome. An ensemble of machine learning models using the super learner algorithm is used to estimate the regressions for treatment and outcome(25,26). Additional methodological details are available in Supplemental Materials. A code tutorial is available at https://github.com/kathoffman/steroids-trial-emulation.

Emulation of model-first approaches common in clinical literature

For contrast with the target trial emulation strategy, we review methodology of papers cited in Chaharom et al.'s(27) meta-analysis, and then analyze the data using study designs common in other corticosteroids for COVID-19 observational research. The data source, outcome, and confounders are the same as the above target trial. Modifications to the cohort and treatment definitions to accommodate the model-first approaches are outlined below.

Cox models using a point treatment

The first approach we explore is a regression for mortality with a point (as opposed to time-varying) treatment variable. The inclusion criteria and time zero are defined as the time of meeting hypoxia criteria, which is the intended indication for corticosteroids. A study design using this analytical approach entails a number of choices, including:

1. Defining a range of time relative to inclusion criteria for a patient to be considered "treated".
2. Deciding whether to exclude patients treated before the inclusion time.
3. Deciding how to handle patients who died during the treatment time window.
4. Deciding how to handle patients treated after the treatment time window.

We fit Cox proportional hazards models using data sets obtained from various design choices, summarized in Table 1. All baseline confounders, time-dependent confounders from day zero, and the corticosteroid exposure are included as variables in the Cox models. The exponentiated

medRxiv preprint doi: https://doi.org/10.1101/2022.05.27.22275037; this version posted June 7, 2022. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted medRxiv a license to display the preprint in perpetuity. It is made available under a CC-BY-NC-ND 4.0 International license .

coefficient for corticosteroids is interpreted as the hazard ratio for corticosteroid exposure within the defined treatment window for moderate-to-severe COVID-19 patients.

Time-varying Cox models

In our second model-first approach, we fit a time-varying Cox model for time to mortality up to 28 days from the day of hospitalization. This model uses our entire cohort and contains all baseline and time-dependent confounders, as well as daily corticosteroid administration. The coefficient for corticosteroids is exponentiated and used as an estimate of the hazard ratio for corticosteroids.

Software

All data were analyzed in R version 4.0.3 with open-source packages ggplot2, gtsummary, survival, survminer, lmtp, and sl3(28–34).

RCT benchmark

Several RCTs have established the effectiveness of corticosteroids in the treatment of moderate-to-severe COVID-19 patients(35–37). The WHO performed a meta-analysis of seven such RCTs and estimated the odds ratio (OR) of mortality to be 0.66 (95% CI (0.53-0.82)(14). We use this estimate, as well as supporting evidence from other RCT meta-analyses(27,38) to benchmark our results.

Results

Target trial emulation

The final cohort includes 3,298 patients of a median age 65 (IQR 53, 77) and 60% males. The median BMI was 27 (IQR 23-31). There were 1033 (31%) patients with DM, 460 (14%) with CAD, 1780 (54%) with hypertension, and 159 (4.8%) with CKD/ESRD. Table 2 shows baseline characteristics of the cohort, overall and stratified by any corticosteroid exposure. There were 1,690 patients who reached the randomization criteria of severe hypoxia and 423 patients who received corticosteroids at any point during follow-up. 699 patients died before 28 days.

medRxiv preprint doi: https://doi.org/10.1101/2022.05.27.22275037; this version posted June 7, 2022. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted medRxiv a license to display the preprint in perpetuity. It is made available under a CC-BY-NC-ND 4.0 International license .

In the target trial emulation analysis, all 3,298 patients who were admitted to the hospital are analyzed. The estimated mortality rate under a hypothetical intervention of no corticosteroids is 32% (95% CI 31-34). The estimated mortality rate under a hypothetical intervention in which corticosteroids are administered for six days upon patients becoming severely hypoxic is 23% (21-24). This yields an estimated mortality reduction of 9.6% (8.8-10.4) if this policy had been implemented.

Model-first approaches

In a subset of 1,690 patients who met severe hypoxia, 72 patients received corticosteroids within one day of hypoxia and 191 patients received corticosteroids within 5 days of hypoxia. There were 18 and 451 patients who died within one and five days of hypoxia without receiving corticosteroids, respectively.

Model A, which defined corticosteroid exposure as anytime during hospitalization, yielded an HR of 0.50 (0.41-0.62). Models B-I, which placed either a one- or five-day limit on corticosteroids treatment from the time of hypoxia, yielded mostly non-significant HRs in both directions (B: 0.95 (0.66-1.37), C: 0.92 (0.63-1.33), D: 0.89 (0.56-1.41), E: 0.66 (0.41-1.04), G: 1.05 (0.77-1.45), H: 1.04 (0.75-1.45)). The exception to this was Model I, which excluded patients who died before five days and estimated the HR to be 0.63 (0.48-0.83). Model F also reached statistical significance, 0.77 (0.60-0.99), and was the result of a 5-day treatment window with no exclusion or censoring variations. The time-varying Cox model (J), yielded an HR of 1.08 (0.80-1.47). HRs for the model-first approaches are summarized in Figure 3.

Discussion

Our research illustrates how a question-first approach can aid in devising an optimal design and choice of estimation procedure for an analysis of observational data. Specifically, we show that using the target trial framework succeeds in recovering the benchmark causal effect obtained in RCTs. Our final estimate that corticosteroids would reduce overall 28-day mortality in a hospitalized cohort by 9.6% is equivalent to an OR of 0.62, which is qualitatively identical to the WHO's estimate of 0.66. Our study design allowed us to create a realistic trial with a meaningful intervention, i.e. randomize patients at hospitalization but do not give corticosteroids unless the

medRxiv preprint doi: https://doi.org/10.1101/2022.05.27.22275037; this version posted June 7, 2022. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted medRxiv a license to display the preprint in perpetuity. It is made available under a CC-BY-NC-ND 4.0 International license .

patient becomes severely hypoxic. Our analysis plan enabled us to flexibly adjust for a large number of potential time-dependent confounders of the treatment, censoring, and outcome mechanisms.

In contrast, the majority of our model-first approaches could not recover the RCT benchmark using the same data source. This finding aligns with other corticosteroids research; a recent meta-analysis containing observational analyses on over 18,000 patients found no overall effect for corticosteroids on mortality (OR 1.12, (0.83–1.50))(27). The task of creating reliable evidence from complex longitudinal data is not an easy one, and many of these studies suffer from flawed designs.

We found most studies in the current observational corticosteroids literature allowed the "treated" group to receive corticosteroids anytime during hospitalization(39–41). This is problematic because it introduces immortal time and biases results towards a protective effect of corticosteroids(42). A few studies did limit the treatment time frame in an effort to diminish immortal time bias. The "grace period" for treatment was handled in various ways, e.g. excluding patients who die prior to a time window after inclusion criteria(43), or excluding patients who receive treatment after the treatment window ends(44,45). Both exclusions are incorrect and may lead to bias and spurious associations(1). An alternative to exclusion is censoring patients at their time of receiving treatment if that time is after the treatment window passes. However, since Cox regression cannot handle time-dependent censoring, this biases results in favor of corticosteroids(1).

In addition to these issues, it is often unclear in the current literature how patients who receive corticosteroids prior to meeting inclusion criteria are handled in the analysis(39–41,46). A related issue is that corticosteroids can, according to RCTs, affect severity of illness (e.g. severe hypoxia). All of the point treatment studies are thus subject to this form of collider bias(47). Although the time-varying treatment approach does not suffer from the same time-alignment biases as the point-exposure design, the time-varying Cox model cannot properly account for time-dependent confounders(1), such as the relationship between intubation, corticosteroids, and mortality. These biases appear in our model-first results; the only study designs which demonstrate a protective effect of corticosteroids in line with the RCT benchmark suffer from extreme immortal time bias through undefined or extended treatment time windows (A, I).

medRxiv preprint doi: https://doi.org/10.1101/2022.05.27.22275037; this version posted June 7, 2022. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted medRxiv a license to display the preprint in perpetuity. It is made available under a CC-BY-NC-ND 4.0 International license .

There are limitations of our study. First, while the study time frame is ideal in terms of corticosteroid experimentation, it includes New York City's initial pandemic surge conditions and rapidly changing clinical practice. We cannot rule out the presence of unmeasured confounding. Second, we did not have the data to look at individual corticosteroid types, making an exact comparison to a specific randomized trial impossible.

Despite these limitations, our study serves as an example in which the current standard for clinical research methods fail to recover the correct treatment effect where a modern causal inference method succeeds. Using observational data to guide clinical practice is possible, but relies on the incorporation of advanced epidemiological and statistical methodology principles. We hope this study and accompanying technical guide encourages the incorporation of these innovative techniques into study designs and statistical analyses of observational data.

Acknowledgements

The authors thank all of the healthcare workers who courageously expanded their roles during the pandemic's surge conditions. This work was made possible through data provided by the Cornell COVID-19 Registry, led by Parag G. Goyal, M.D., Justin Choi, M.D., Laura Pinheiro, Ph.D., and Monika Safford, M.D., of Weill Cornell Medicine. The authors also thank the contributions to this work of the Architecture for Research Computing in Health team.

Funding: This research study was not funded by any grants. **Conflicts of interest:** Dr. Edward Schenck is supported by NHLBI HL151876 and reports consulting for Axle Informatics regarding Coronavirus vaccine clinical trial through NIAID outside of the current work. Dr. Michael Satlin is supported by research grants from Allergan, Merck, BioFire Diagnostics, and SNIPRBiome and reports consulting payments from Shionogi outside of the current work.

medRxiv preprint doi: https://doi.org/10.1101/2022.05.27.22275037; this version posted June 7, 2022. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted medRxiv a license to display the preprint in perpetuity. It is made available under a CC-BY-NC-ND 4.0 International license .

References

[1] Robins JM Hernan MA. *Causal Inference: What If*. Boca Raton: Chapman & Hall, 2020.

[2] Miguel A. Hernan.´ Methods of public health research — strengthening causal inference from observational data. *New England Journal of Medicine*, 385(15):1345–1348, 2021. doi: 10.1056/NEJMp2113319. URL https://doi.org/10.1056/NEJMp2113319. PMID: 34596980.

[3] Steven S. Henley, Richard M. Golden, and T. Michael Kashner. Statistical modeling methods: challenges and strategies. *Biostatistics & Epidemiology*, 4(1):105–139, 2020. doi: 10.1080/ 24709360.2019.1618653. URL https://doi.org/10.1080/24709360.2019.1618653.

[4] Mohammad Ali Mansournia, Mahyar Etminan, Goodarz Danaei, Jay S Kaufman, and Gary Collins. Handling time varying confounding in observational research. *BMJ*, 359, 2017. ISSN 0959-8138. doi: 10.1136/bmj.j4587. URL https://www.bmj.com/content/359/bmj.j4587.

[5] Leo Breiman. Statistical modeling: The two cultures (with comments and a rejoinder by the author). *Statistical science*, 16(3):199–231, 2001.

[6] Miguel A. Hernan. The Hazards of Hazard Ratios.´ *Epidemiology*, 21(1):13–15, 01 2010. doi: 10.1097/EDE.0b013e3181c1ea43.

[7] Mats J. Stensrud and Miguel A. Hernan. Why Test for Proportional Hazards?´ *JAMA*, 323 (14):1401–1402, 04 2020. ISSN 0098-7484. doi: 10.1001/jama.2020.1267. URL https://doi.org/10.1001/jama.2020.1267.

[8] Daniel Westreich and Sander Greenland. The Table 2 Fallacy: Presenting and Interpreting Confounder and Modifier Coefficients. *American Journal of Epidemiology*, 177(4):292–298, 01 2013. ISSN 0002-9262. doi: 10.1093/aje/kws412. URL https://doi.org/10.1093/aje/kws412.

[9] Gary Smith. Step away from stepwise. *Journal of Big Data*, 5(1):1–12, 2018.

[10] Miguel A. Hernan and James M. Robins. Using Big Data to Emulate a Target Trial When´ a Randomized Trial Is Not Available. *American Journal of Epidemiology*, 183(8):758–764, 03 2016. ISSN 0002-9262. doi: 10.1093/aje/kwv254. URL https://doi.org/10.1093/aje/kwv254.

[11] Maya L. Petersen and Mark J. van der Laan. Causal models and learning from data: Integrating causal modeling and statistical estimation. *Epidemiology*, 25(3):418–426, 2014. ISSN 10443983. URL http://www.jstor.org/stable/24759134.

[12] Jeremy A Labrecque and Sonja A Swanson. Target trial emulation: teaching epidemiology and beyond. *European journal of epidemiology*, 32(6):473–475, 2017.

[13] Miguel A. Hernan, Brian C. Sauer, Sonia Hern´ andez-D´ ´ıaz, Robert Platt, and Ian Shrier. Specifying a target trial prevents immortal time bias and other self-inflicted injuries in

medRxiv preprint doi: https://doi.org/10.1101/2022.05.27.22275037; this version posted June 7, 2022. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted medRxiv a license to display the preprint in perpetuity. It is made available under a CC-BY-NC-ND 4.0 International license .

observational analyses. *Journal of Clinical Epidemiology*, 79:70–75, 2016. ISSN 0895-4356. doi: https://doi.org/10.1016/j.jclinepi.2016.04.014. URL https://www.sciencedirect.com/ science/article/pii/S0895435616301366.

[14] Jonathan AC Sterne, Srinivas Murthy, Janet V Diaz, Arthur S Slutsky, Jesus Villar, Derek C´ Angus, Djillali Annane, Luciano Cesar Pontes Azevedo, Otavio Berwanger, Alexandre B Cavalcanti, et al. Association between administration of systemic corticosteroids and mortality among critically ill patients with covid-19: a meta-analysis. *Jama*, 324(13):1330–1341, 2020.

[15] Steroid conversion calculator. URL https://www.mdcalc.com/ steroid-conversion-calculator.

[16] Parag Goyal, Justin J Choi, Laura C Pinheiro, Edward J Schenck, Ruijun Chen, Assem Jabri, Michael J Satlin, Thomas R Campion Jr, Musarrat Nahid, Joanna B Ringel, et al. Clinical characteristics of covid-19 in new york city. *New England Journal of Medicine*, 382(24): 2372–2374, 2020.

[17] Edward J Schenck, Katherine L Hoffman, Marika Cusick, Joseph Kabariti, Evan T Sholle, and Thomas R Campion Jr. Critical care database for advanced research (cedar): An automated method to support intensive care units with electronic health record data. *Journal of Biomedical Informatics*, 118:103789, 2021.

[18] Bibhas Chakraborty and EE Moodie. Statistical methods for dynamic treatment regimes. *Springer-Verlag. doi*, 10:978–1, 2013.

[19] James Robins. A new approach to causal inference in mortality studies with a sustained exposure period—application to control of the healthy worker survivor effect. *Mathematical Modelling*, 7(9):1393–1512, 1986. ISSN 0270-0255. doi: https://doi.org/10.1016/ 0270-0255(86)90088-6. URL https://www.sciencedirect.com/science/article/ pii/0270025586900886.

[20] Alexander R Luedtke, Oleg Sofrygin, Mark J van der Laan, and Marco Carone. Sequential double robustness in right-censored longitudinal models. *arXiv preprint arXiv:1705.02459*, 2017.

[21] Ivan D´ ´ıaz, Nicholas Williams, Katherine L. Hoffman, and Edward J. Schenck. Nonparametric causal effects based on longitudinal modified treatment policies. *Journal of the American Statistical Association*, 0(0):1–16, 2021. doi: 10.1080/01621459.2021.1955691. URL https://doi.org/10.1080/01621459.2021.1955691.

[22] Mark J. van der Laan and Sherri Rose. *Targeted Learning in Data Science: Causal Inference for Complex Longitudinal Studies*. Springer Publishing Company, Incorporated, 1st edition, 2018. ISBN 3319653032.

[23] Victor Chernozhukov, Denis Chetverikov, Mert Demirer, Esther Duflo, Christian Hansen, Whitney Newey, and James Robins. Double/debiased machine learning for treatment and

medRxiv preprint doi: https://doi.org/10.1101/2022.05.27.22275037; this version posted June 7, 2022. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted medRxiv a license to display the preprint in perpetuity. It is made available under a CC-BY-NC-ND 4.0 International license .

structural parameters. *The Econometrics Journal*, 21(1):C1–C68, 01 2018. ISSN 1368-4221. doi: 10.1111/ectj.12097. URL https://doi.org/10.1111/ectj.12097.

[24] T. A. Gerds and M. Schumacher. On functional misspecification of covariates in the Cox regression model. *Biometrika*, 88(2):572–580, 06 2001. ISSN 0006-3444. doi: 10.1093/ biomet/88.2.572. URL https://doi.org/10.1093/biomet/88.2.572.

[25] Leo Breiman. Stacked regressions. *Machine learning*, 24(1):49–64, 1996.

[26] Mark J. van der Laan, Eric C Polley, and Alan E. Hubbard. Super learner. *Statistical Applications in Genetics and Molecular Biology*, 6(1), 2007. doi: doi:10.2202/1544-6115.1309. URL https://doi.org/10.2202/1544-6115.1309.

[27] Faegheh Ebrahimi Chaharom, Leili Pourafkari, Ali Asghar Ebrahimi Chaharom, and Nader D Nader. Effects of corticosteroids on covid-19 patients: A systematic review and meta-analysis on clinical outcomes. *Pulmonary pharmacology & therapeutics*, page 102107, 2021.

[28] Hadley Wickham. *ggplot2: Elegant Graphics for Data Analysis*. Springer-Verlag New York, 2016. ISBN 978-3-319-24277-4. URL https://ggplot2.tidyverse.org.

[29] Daniel D. Sjoberg, Karissa Whiting, Michael Curry, Jessica A. Lavery, and Joseph Larmarange. Reproducible summary tables with the gtsummary package. *The R Journal*, 13:570–580, 2021. doi: 10.32614/RJ-2021-053. URL https://doi.org/10.32614/ RJ-2021-053.

[30] Terry M Therneau. *A Package for Survival Analysis in R*, 2021. URL https://CRAN. R-project.org/package=survival. R package version 3.2-13.

[31] Terry M. Therneau and Patricia M. Grambsch. *Modeling Survival Data: Extending the Cox Model*. Springer, New York, 2000. ISBN 0-387-98784-3.

[32] Alboukadel Kassambara, Marcin Kosinski, and Przemyslaw Biecek. *survminer: Drawing Survival Curves using 'ggplot2'*, 2021. URL https://CRAN.R-project.org/package= survminer. R package version 0.4.9.

[33] Nicholas T Williams, Katherine L Hoffman, and Ivan D´ ´ıaz. lmtp: Non-parametric causal effects of feasible interventions based on modified treatment policies, 2022. URL https: //github.com/hoffmakl/lmtp_comprisks. R package version 1.0.1.

[34] Jeremy R Coyle, Nima S Hejazi, Ivana Malenica, and Oleg Sofrygin. *sl3: Pipelines for Machine Learning and Super Learning*, 2022. URL https://github.com/tlverse/sl3. R package version 1.4.2.

[35] Bruno Martins Tomazini, Israel Silva Maia, Flavia Regina Bueno, Maria Vitoria Aparecida Oliveira Silva, Franca Pellison Baldassare, Eduardo Leite Vieira Costa, Ricardo Antonio Bonifacio Moura, Michele Ouriques Honorato, Andr´ e Nathan Costa, Alexandre Biasi´

medRxiv preprint doi: https://doi.org/10.1101/2022.05.27.22275037; this version posted June 7, 2022. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted medRxiv a license to display the preprint in perpetuity. It is made available under a CC-BY-NC-ND 4.0 International license .

Cavalcanti, et al. Covid-19-associated ards treated with dexamethasone (codex): study design and rationale for a randomized trial. *Revista Brasileira de terapia intensiva*, 32:354–362, 2020.

[36] Maryam Edalatifard, Maryam Akhtari, Mohammadreza Salehi, Zohre Naderi, Ahmadreza Jamshidi, Shayan Mostafaei, Seyed Reza Najafizadeh, Elham Farhadi, Nooshin Jalili, Masoud Esfahani, et al. Intravenous methylprednisolone pulse as a treatment for hospitalised severe covid-19 patients: results from a randomised controlled clinical trial. *European Respiratory Journal*, 56(6), 2020.

[37] Luis Corral-Gudino, Alberto Bahamonde, Francisco Arnaiz-Revillas, Julia Gomez-Barquero,´ Jesica Abad´ıa-Otero, Carmen Garc´ıa-Ibarbia, V´ıctor Mora, Ana Cerezo-Hernandez, Jos´ e L´ Hernandez, Graciela L´ opez-Mu´ n˜´ız, et al. Methylprednisolone in adults hospitalized with covid-19 pneumonia. *Wiener klinische Wochenschrift*, 133(7):303–311, 2021.

[38] Carina Wagner, Mirko Griesel, Agata Mikolajewska, Anika Mueller, Monika Nothacker, Karoline Kley, Maria-Inti Metzendorf, Anna-Lena Fischer, Marco Kopp, Miriam Stegemann, et al. Systemic corticosteroids for the treatment of covid-19. *Cochrane Database of Systematic Reviews*, (8), 2021.

[39] Ana Fernandez-Cruz, Bel´ en Ruiz-Antor´ an, Ana Mu´ noz-G˜ omez, Ar´ anzazu Sancho-L´ opez,´ Patricia Mills-Sanchez,´ Gustavo Adolfo Centeno-Soto, Silvia Blanco-Alonso, Laura Javaloyes-Garachana, Amy Galan-G´ omez,´ Angela Valencia-Alijo, et al. A retrospective con-´ trolled cohort study of the impact of glucocorticoid treatment in sars-cov-2 infection mortality. *Antimicrobial agents and chemotherapy*, 64(9):e01168–20, 2020.

[40] Brian C Nelson, Justin Laracy, Sherif Shoucri, Donald Dietz, Jason Zucker, Nina Patel, Magdalena E Sobieszczyk, Christine J Kubin, and Angela Gomez-Simmonds. Clinical outcomes associated with methylprednisolone in mechanically ventilated patients with covid-19. *Clinical Infectious Diseases*, 72(9):e367–e372, 2021.

[41] Jiao Liu, Sheng Zhang, Xuan Dong, Zhongyi Li, Qianghong Xu, Huibin Feng, Jing Cai, Sisi Huang, Jun Guo, Lidi Zhang, et al. Corticosteroid treatment in severe covid-19 patients with acute respiratory distress syndrome. *The Journal of clinical investigation*, 130(12):6417– 6428, 2020.

[42] Linda E Levesque, James A Hanley, Abbas Kezouh, and Samy Suissa. Problem of immortal´ time bias in cohort studies: example using statins for preventing progression of diabetes. *Bmj*, 340, 2010.

[43] Michele Bartoletti, Lorenzo Marconi, Luigia Scudeller, Livia Pancaldi, Sara Tedeschi, Maddalena Giannella, Matteo Rinaldi, Linda Bussini, Ilaria Valentini, Anna Filomena Ferravante, et al. Efficacy of corticosteroid treatment for hospitalized patients with severe covid-19: a multicentre study. *Clinical Microbiology and Infection*, 27(1):105–111, 2021.

medRxiv preprint doi: https://doi.org/10.1101/2022.05.27.22275037; this version posted June 7, 2022. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted medRxiv a license to display the preprint in perpetuity. It is made available under a CC-BY-NC-ND 4.0 International license .

[44] Jesus Rodr´ ´ıguez-Bano, Jer˜ onimo Pach´ on, Jordi Carratal´ a, Pablo Ryan, Inmaculada Jarr` ´ın, Mar´ıa Yllescas, Jose Ram´ on Arribas, Juan Berenguer, Esther Aznar Mu´ noz, Pedro Gil Di-˜ vasson, et al. Treatment with tocilizumab or corticosteroids for covid-19 patients with hyperinflammatory state: a multicentre cohort study (sam-covid-19). *Clinical Microbiology and Infection*, 27(2):244–252, 2021.

[45] Ivan Cusacovich, Alvaro Aparisi, Miguel Marcos, Cristina Ybarra-Falc´ on, Carolina Iglesias-´ Echevarria, Maria Lopez-Veloso, Julio Barraza-Vengoechea, Carlos Duenas, Santiago Anto-˜ nio Juarros Mart´ınez, Beatriz Rodr´ıguez-Alonso, et al. Corticosteroid pulses for hospitalized patients with covid-19: effects on mortality. *Mediators of inflammation*, 2021, 2021.

[46] Cecilia Tortajada, Enrique Colomer, Juan C Andreu-Ballester, Ana Esparcia, Carmina Oltra, and Juan Flores. Corticosteroids for covid-19 patients requiring oxygen support? yes, but not for everyone: effect of corticosteroids on mortality and intensive care unit admission in patients with covid-19 according to patients' oxygen requirements. *Journal of Medical Virology*, 93(3):1817–1823, 2021.

[47] Gareth J Griffith, Tim T Morris, Matthew J Tudball, Annie Herbert, Giulia Mancano, Lindsey Pike, Gemma C Sharp, Jonathan Sterne, Tom M Palmer, George Davey Smith, et al. Collider bias undermines our understanding of covid-19 disease risk and severity. *Nature communications*, 11(1):1–12, 2020.

[48] Heejung Bang and James M Robins. Doubly robust estimation in missing data and causal inference models. *Biometrics*, 61(4):962–973, 2005.

medRxiv preprint doi: https://doi.org/10.1101/2022.05.27.22275037; this version posted June 7, 2022. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted medRxiv a license to display the preprint in perpetuity. It is made available under a CC-BY-NC-ND 4.0 International license .

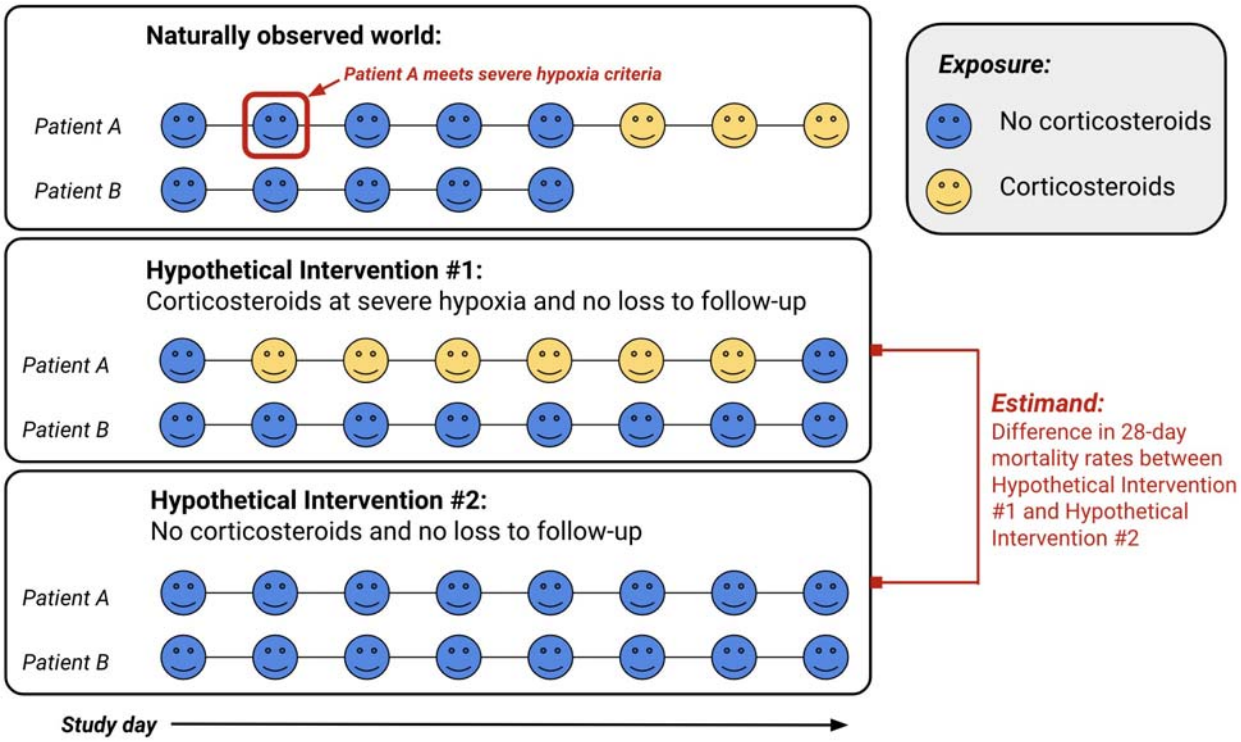

Figure 1: Illustrated example of two patients under the two hypothetical interventions of our target trial emulation. Patient A reaches severe hypoxia criteria at study day 2 and is followed the entire study duration. Patient B never reaches severe hypoxia criteria and is lost to follow up after five study days. Under the dynamic corticosteroids intervention (Intervention #1), Patient A receives 6 days of corticosteroids, and under Intervention #2 they receive no steroids. Patient B does not receive corticosteroids under either intervention strategy, however, in both hypothetical worlds they are observed for the entire study duration.

medRxiv preprint doi: https://doi.org/10.1101/2022.05.27.22275037; this version posted June 7, 2022. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted medRxiv a license to display the preprint in perpetuity. It is made available under a CC-BY-NC-ND 4.0 International license .

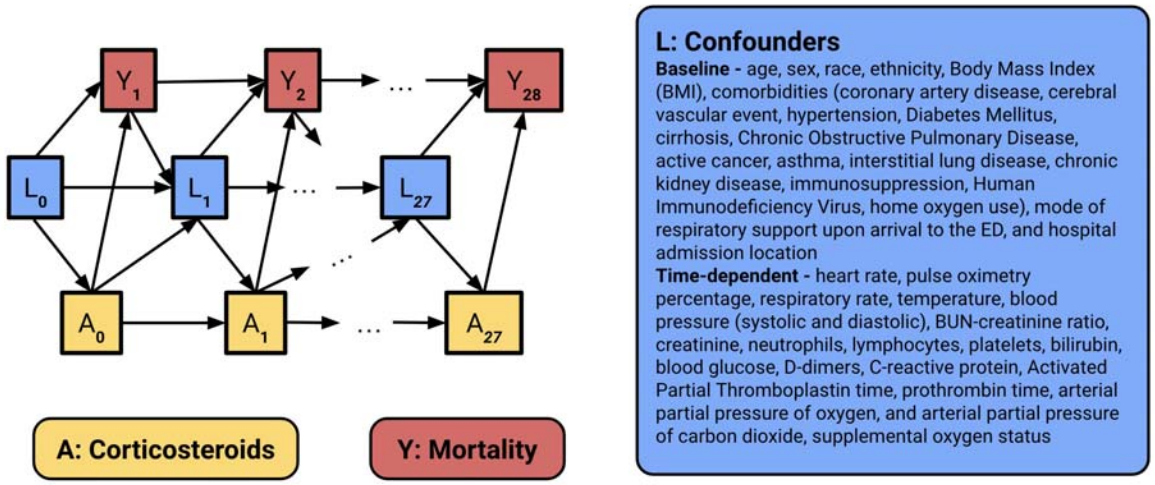

Figure 2: Directed Acyclic Graph (DAG) showing the relationship between confounders L_t, corticosteroid exposure A_t, and mortality Y_t. Baseline confounders are included in L_0.

medRxiv preprint doi: https://doi.org/10.1101/2022.05.27.22275037; this version posted June 7, 2022. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted medRxiv a license to display the preprint in perpetuity. It is made available under a CC-BY-NC-ND 4.0 International license .

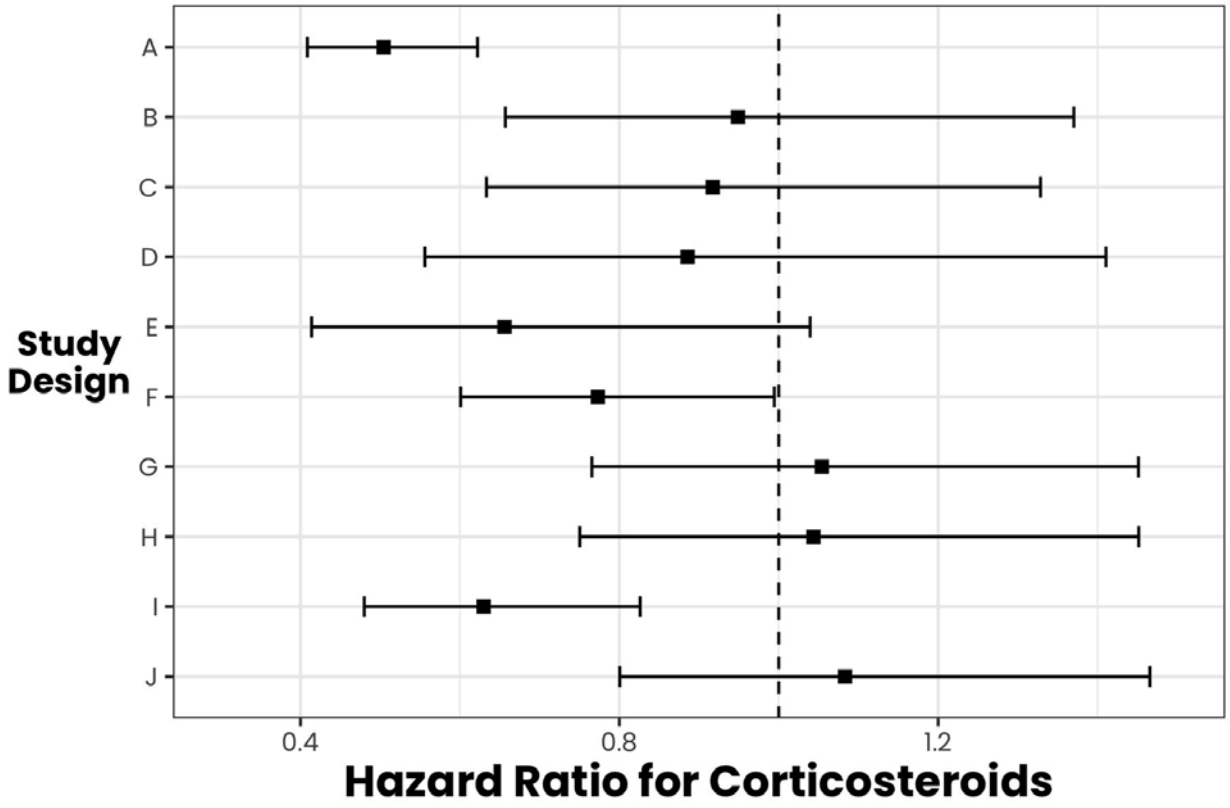

Figure 3: Forest plot of model-first results. Study designs A-J correspond to Table 1's specifications. Hazard ratio estimates (95% confidence interval) for each model are as follows: Model A: 0.504 (0.409-0.622), Model B: 0.949 (0.657-1.37), Model C: 0.917 (0.633-1.328), Model D: 0.886 (0.556-1.411), Model E: 0.656 (0.414-1.039), Model F: 0.773 (0.601-0.994), Model G: 1.054 (0.766-1.451), Model H: 1.044 (0.75-1.452), Model I: 0.63 (0.48-0.826), and Model J: 1.083 (0.801-1.466).

medRxiv preprint doi: https://doi.org/10.1101/2022.05.27.22275037; this version posted June 7, 2022. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted medRxiv a license to display the preprint in perpetuity. It is made available under a CC-BY-NC-ND 4.0 International license .

Model	Study Design
A	Corticosteroid exposure defined as anytime during the course of hospitalization. All patients satisfying inclusion criteria are included in the analysis and time to event is defined as time from hypoxia to death.
B	Corticosteroid exposure defined as any administration up to one day after meeting hypoxia criteria. All patients satisfying inclusion criteria are included in the analysis and time to event is defined as time from hypoxia to death.
C	Corticosteroid exposure defined as any administration up to one day after meeting hypoxia criteria. Patients who died during this time window are excluded. Patients who receive corticosteroids after the time window are included in the control group.
D	Corticosteroid exposure defined as any administration up to one day after meeting hypoxia criteria. Patients who died during this time window are excluded. Patients who receive corticosteroids before hypoxia are excluded. Patients who receive corticosteroids after the time window are included in the control group.
E	Corticosteroid exposure defined as any administration up to one day after meeting hypoxia criteria. Patients who receive corticosteroids before hypoxia are excluded. Patients who receive corticosteroids after the one-day time window passes are censored at the time of corticosteroids receipt.
F	Corticosteroid exposure defined as any administration up to five days after meeting hypoxia criteria. All patients satisfying inclusion criteria are included in the analysis and time to event is defined as time from hypoxia to death.
G	Corticosteroid exposure defined as any administration up to five days after meeting hypoxia criteria. Patients who died during this time window are excluded. Patients who receive corticosteroids after the time window are included in the control group.
H	Corticosteroid exposure defined as any administration up to five days after meeting hypoxia criteria. Patients who died during this time window are excluded. Patients who receive corticosteroids before hypoxia are excluded. Patients who receive corticosteroids after the time window are included in the control group.
I	Corticosteroid exposure defined as any administration up to five days after meeting hypoxia criteria. Patients who receive

medRxiv preprint doi: https://doi.org/10.1101/2022.05.27.22275037; this version posted June 7, 2022. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted medRxiv a license to display the preprint in perpetuity. It is made available under a CC-BY-NC-ND 4.0 International license .

	corticosteroids before hypoxia are excluded. Patients who receive corticosteroids after the one-day time window passes are censored at the time of corticosteroids receipt.
J	Corticosteroid exposure is allowed to be a time-varying covariate beginning at the time of hospitalization.

Table 1: Study design specifications for the model-first approaches.

medRxiv preprint doi: https://doi.org/10.1101/2022.05.27.22275037; this version posted June 7, 2022. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted medRxiv a license to display the preprint in perpetuity. It is made available under a CC-BY-NC-ND 4.0 International license .

Characteristic[1]	Overall [N=3,298]	Corticosteroid Never [N=2,875]	Corticosteroids Ever [N=423]
Age	65 (53, 77)	65 (52, 77)	67 (58, 75)
Male	1,970 (60%)	1,697 (59%)	273 (65%)
Race			
Asian	602 (18%)	517 (18%)	85 (20%)
Black	399 (12%)	352 (12%)	47 (11%)
White	938 (28%)	818 (28%)	120 (28%)
Other	1,141 (35%)	1,009 (35%)	132 (31%)
Unknown	218 (6.6%)	179 (6.2%)	39 (9.2%)
BMI	27 (23, 31)	27 (23, 31)	28 (24, 32)
Home supplemental oxygen	312 (9.5%)	286 (9.9%)	26 (6.1%)
Coronary Artery Disease	460 (14%)	402 (14%)	58 (14%)
Diabetes Mellitus	1,033 (31%)	891 (31%)	142 (34%)
Hypertension	1,780 (54%)	1,544 (54%)	236 (56%)
Cerebral Vascular Event	225 (6.8%)	193 (6.7%)	32 (7.6%)
Cirrhosis	35 (1.1%)	30 (1.0%)	5 (1.2%)
CKD/ESRD	159 (4.8%)	146 (5.1%)	13 (3.1%)
Asthma	180 (5.5%)	145 (5.0%)	35 (8.3%)
COPD	134 (4.1%)	100 (3.5%)	34 (8.0%)
Active cancer	136 (4.1%)	118 (4.1%)	18 (4.3%)
Immunosuppressed	51 (1.5%)	44 (1.5%)	7 (1.7%)
ILD	5 (0.2%)	3 (0.1%)	2 (0.5%)
HIV	35 (1.1%)	33 (1.1%)	2 (0.5%)
Active smoker	104 (3.2%)	93 (3.2%)	11 (2.6%)
Former smoker	543 (16%)	442 (15%)	101 (24%)
***Outcome:* 28-day mortality**	699 (21%)	574 (20%)	125 (30%)

[1]All continuous variables are reported as median (interquartile range) and categorical variables are n (%). Abbreviations: BMII=Body Mass Index, CKD=Chronic Kidney Disease, ESRD=End Stage Renal Disease, COPD=Chronic Obstructive Pulmonary Disease, ILD=Interstitial Lung Disease, HIV=Human Immunodeficiency Virus

Table 2: Demographics and outcome for study cohort, overall and stratified by any corticosteroid exposure.